
UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT

MID CONTINENT STEEL & WIRE, INC.,

Plaintiff-Appellee,

v.

UNITED STATES,

Defendant-Appellee,

and

PT ENTERPRISE INC., PRO-TEAM COIL NAIL ENTERPRISE INC.,
UNICATCH INDUSTRIAL CO., LTD., WTA INTERNATIONAL CO., LTD.,
ZON MON CO., LTD., HOR LIANG INDUSTRIAL CORPORATION,

PRESIDENT INDUSTRIAL INC., and
LIANG CHYUAN INDUSTRIAL CO., LTD.,

Defendants-Appellants.

Appeal From The United States Court Of International Trade, Court
No. 15-00213, Slip Op. 24-15 (Entered February 12, 2024), Judge Claire Kelly

BRIEF FOR DEFENDANT-APPELLEE, UNITED STATES

BRIAN M. BOYNTON

Principal Deputy Assistant Attorney General

PATRICIA M. MCCARTHY

Director

MIKKI COTTET

Senior Trial Counsel

Commercial Litigation Branch

Civil Division

U.S. Department of Justice

P.O. Box 480, Ben Franklin Station

Washington, D.C. 20044

Tel: (202) 307-0962

Email: Mikki.Cottet@usdoj.gov

Attorneys for Defendant-Appellee,

United States

OF COUNSEL:

VANIA WANG

Attorney

U.S. Department of Commerce

Office of the Chief Counsel

for Trade Enforcement and

Compliance

August 15, 2024

TABLE OF CONTENTS

STATEMENT OF RELATED CASES	1
STATEMENT OF THE ISSUE.....	3
STATEMENT OF THE CASE.....	4
I. <i>Mid Continent V</i> Background.....	5
A. Initial Administrative Proceeding.....	5
B. Judicial Challenges To Commerce’s Determinations	7
II. Post- <i>Mid Continent V</i> Proceedings.....	9
SUMMARY OF THE ARGUMENT	12
ARGUMENT	14
I. Standard Of Review	14
II. Commerce’s Determination To Use A Simple Average For The Denominator Of The Cohen’s <i>d</i> Test Is Reasonable	16
A. Commerce’s Differential Pricing Analysis And Cohen’s <i>d</i> Test	16
B. The Principle That Commerce Derived From The Academic Literature Leads To A Logical Conclusion That Use Of A Simple Average To Calculate The Denominator Of Commerce’s Cohen’s <i>d</i> Coefficient Is Reasonable.....	20
1. The Standard Deviations Of Two Different Sized Full Populations Are Equally Reliable Just As The Standard Deviations Of Two Samples Of Equal Size Are Equally Reliable.....	20
2. PT’s And Canada’s Challenges To Commerce’s Reliability Analysis As Supporting The Use Of A Simple Average Lack Merit	25

C. Commerce Reasonably Determined That Use Of A Single Standard Deviation Is Not Appropriate.....	35
D. PT’s Arguments That Commerce Should Use Weight Averaging Are Unavailing	38
III. The New Issues And Arguments Raised By Canada Should Be Rejected Because They Stray Beyond The Sole Question Before The Court And The Arguments Raised In PT’s Opening Brief And Are Supported By Non-Record Evidence	46
A. Canada’s New Issues And Arguments Should Be Deemed Waived Because PT’s Brief Does Not Raise Or Otherwise Include Them.....	48
B. Canada’s New Arguments And Non-Record Supporting Materials Should Be Stricken Or Disregarded Because They Were Not Before Commerce During The Administrative Proceedings	50
C. Canada’s New Arguments Are Unavailing	54
CONCLUSION	60

TABLE OF AUTHORITIES

Cases	Page(s)
<i>Ad Hoc Shrimp Trade Action Comm. v. United States</i> , 791 F. Supp. 2d 1327 (Ct. Int’l Trade 2011).....	52
<i>Amoco Oil Co. v. United States</i> , 234 F.3d 1374 (Fed. Cir. 2000)	48
<i>Apex Frozen Foods Private Ltd. v. United States</i> , 144 F. Supp. 3d 1308 (Ct. Int’l Trade 2016), <i>aff’d</i> , 862 F.3d 1337, 1344 (Fed. Cir. 2017).....	16
<i>Arizona v. California</i> , 460 U.S. 605 (1983)	49
<i>Axiom Res. Mgmt., Inc. v. United States</i> , 564 F.3d 1374 (Fed. Cir. 2009)	52
<i>Borusan Mannesmann Boru Sanayi Ve Ticaret A.S. v. American Cast Iron Pipe Company</i> , 5 F.4th 1367 (Fed. Cir. 2021).....	15, 28
<i>Camp v. Pitts</i> , 411 U.S. 138 (1973)	51
<i>Corus Staal BV v. United States</i> , 502 F.3d 1370 (Fed. Cir. 2007)	51
<i>Fujitsu Gen. Ltd. v. United States</i> , 88 F.3d 1034 (Fed. Cir. 1996)	14
<i>Garg Tube Export LLP v. United States</i> , Court No. 21-00169 (Ct. Int’l Trade).....	3
<i>Government of Canada, et al. v. United States</i> , Court No. 23-00187 (Ct. Int’l Trade).....	3
<i>Home Prods. Int’l Inc. v. United States</i> , 633 F.3d 1369 (Fed. Cir. 2011)	51

<i>Koyo Seiko Co. v. United States</i> , 95 F.3d 1094 (Fed. Cir. 1996)	49
<i>Maquilacero S.A. de C.V. et al. v. United States</i> , Court No. 23-00091 (Ct. Int’l Trade)	3
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 31 F.4th 1367 (Fed. Cir. 2022)	<i>passim</i>
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 219 F. Supp. 3d 1326 (Ct. Int’l Trade 2017)	2, 7
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 273 F. Supp. 3d 1161 (Ct. Int’l Trade 2017)	2, 6, 7
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 495 F. Supp. 3d 1298 (Ct. Int’l Trade 2021)	2, 8
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 628 F. Supp. 3d 1316- (Ct. Int’l Trade 2023)	<i>passim</i>
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 680 F. Supp. 3d 1346 (Ct. Int’l Trade 2024)	<i>passim</i>
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , 940 F.3d 662 (Fed. Cir. 2019)	<i>passim</i>
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , Court Nos. 2018-1229, 2018-1251 (Fed. Cir.)	2
<i>Mid Continent Steel & Wire, Inc. v. United States</i> , Court No. 2021-1747 (Fed. Cir. Sept. 3, 2021)	2, 46
<i>Mobility Workx, LLC v. Unified Patents, LLC</i> , 15 F. 4th 1146 (Fed. Cir. 2021)	46
<i>Murakami v. United States</i> , 46 Fed. Cl. 731 (2000)	52

<i>Nan Ya Plastics Corp. v. United States</i> , 810 F.3d 1333 (Fed. Cir. 2016)	48
<i>Novosteel SA v. United States</i> , 284 F.3d 1261 (Fed. Cir. 2002)	14
<i>Resolute FP Canada Inc. v. United States</i> , Court No. 23-00095 (Ct. Int’l Trade)	3
<i>Robinson v. McDonough</i> , No. 2021-1784, 2022 WL 499845 (Fed. Cir. Feb. 18, 2022).....	49
<i>Sage Prods., Inc. v. Devon Indus., Inc.</i> , 126 F.3d 1420 (Fed. Cir. 1997)	48
<i>SEC v. Chenery Corp.</i> , 332 U.S. 194 (1947)	15
<i>SmithKline Beecham Corp. v. Apotex Corp.</i> , 439 F.3d 1312 (Fed. Cir. 2006)	49
<i>Stupp Corp. v. United States</i> , 5 F.4th 1341 (Fed. Cir. 2021)	15, 52
<i>Suel v. Sec’y of Health & Hum. Servs.</i> , 192 F.3d 981 (Fed. Cir. 1999)	49
<i>Tri Union Frozen Prod., Inc. v. United States</i> , 163 F. Supp. 3d 1255 (Ct. Int’l Trade 2016).....	52
<i>Tri Union Frozen Prods. v. United States</i> , 161 F. Supp. 3d 1333 (Ct. Int’l Trade 2016).....	54
<i>United States v. L.A. Tucker Truck Lines, Inc.</i> , 344 U.S. 33 (1952)	50

Statutes

19 U.S.C. § 1677(35)	16, 42
19 U.S.C. § 1677f-1(d).....	<i>passim</i>
19 U.S.C. § 1516a(b)(1).....	14, 51

Rules

Fed. Cir. R. 47.5	1
Fed. Cir. R. 28(b)	4
Fed. R. of Evidence 201(b)	54

Regulations

19 C.F.R. § 351.102(b)(21).....	52
---------------------------------	----

Administrative Determinations

<i>Certain Steel Nails from India, the Republic of Korea, Malaysia, the Sultanate of Oman, Taiwan, the Republic of Turkey, and the Socialist Republic of Vietnam,</i> 79 Fed. Reg. 36,019 (Dep’t of Commerce June 25, 2014)	5
<i>Certain Steel Nails From the Republic of Korea, Malaysia, the Sultanate of Oman, Taiwan, and the Socialist Republic of Vietnam,</i> 80 Fed. Reg. 39,994 (Dep’t of Commerce July 13, 2015)	4, 6
<i>Certain Steel Nails from Taiwan,</i> 79 Fed. Reg. 78,053 (Dep’t of Commerce Dec. 29, 2014)	5
<i>Certain Steel Nails from Taiwan,</i> 80 Fed. Reg. 28,959 (Dep’t of Commerce May 20, 2015)	2, 4, 5

<i>Differential Pricing Analysis; Request for Comments,</i> 79 Fed. Reg. 26,720, 26,722 (Dep’t of Commerce May 9, 2014)	17
--	----

Other Authorities

CDC, Data Table of Stature-for-age Charts, available at https://www.cdc.gov/growthcharts/html_charts/statage.htm#females (last updated Aug. 23, 2001)	54
--	----

J.P. Guilford, <i>Fundamental Statistics in Psychology and Education</i> (1st ed. 1942)	53
--	----

Jacob Cohen, <i>Statistical Power Analysis for the Behavioral Sciences</i> (2d ed. 1988)	passim
---	--------

Jean-Christophe Goulet-Pelletier & Denis Cousineau, <i>A Review of Effect Sizes and Their Confidence Intervals, Part I: The Cohen’s d</i> <i>Family</i> , 14 <i>The Quantitative Methods for Psychology</i> , 242 (2014)	55, 57
--	--------

Johnson Ching-Hong Li, <i>Effect size measures in a two-independent-samples case with nonnormal and</i> <i>nonhomogeneous data</i> , 48 <i>Behavioral Research</i> 1560, 1561 (2016)	53
--	----

Jum C. Nunnally, <i>Psychometric Theory — 25 Years Ago and Now at 10</i> (Nov. 1975)	53
---	----

Larry V. Hedges, <i>Review and Analysis of the Cohen’s d Test as Used in the U.S. Department of</i> <i>Commerce’s Differential Pricing Methodology</i> (Jan. 2, 2024)	53
---	----

Paul D. Ellis, <i>The Essential Guide to Effect Sizes</i> (2010)	7
---	---

R.E. Deakin, D.G. Kildea, <i>A Note on Standard Deviation and RMS</i> , 44 <i>The Australian Surveyor</i> 74 (June 1999)	53
--	----

Robert Coe, <i>It’s The Effect Size, Stupid: What Effect Size Is And Why It Is Important</i> , Paper	
---	--

presented at the Annual Conf. of British Educational Research Ass'n (Sept. 2002)	8, 36, 37, 58
Steven Wachs, <i>What Is a Standard Deviation and How Do I Compute It?</i> (2009)	53
Xiang Wan, Wenqian Wang, Jiming Liu, Tiejun Tong, <i>Estimating the Sample Mean and Standard Deviation from the Sample Size, Median, Range and/or Interquartile Range</i> , 14:135 BioMed Central Medical Research Methodology (2014)	53

No. 2024-1556

IN THE UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT

MID CONTINENT STEEL & WIRE, INC.
Plaintiff-Appellee,

v.

UNITED STATES,
Defendant-Appellee,

and

PT ENTERPRISE INC.,
PRO-TEAM COIL NAIL ENTERPRISE INC.,
UNICATCH INDUSTRIAL CO., LTD.,
WTA INTERNATIONAL CO., LTD.,
ZON MON CO., LTD.,
HOR LIANG INDUSTRIAL CORPORATION,
PRESIDENT INDUSTRIAL INC., and
LIANG CHYUAN INDUSTRIAL CO., LTD.,
Defendants-Appellants.

Appeal from the United States Court of International Trade
in Case No. 15-00213, Slip Op. 24-15, Judge Claire Kelly

BRIEF OF DEFENDANT-APPELLEE UNITED STATES

STATEMENT OF RELATED CASES

In accordance with Federal Circuit Rule 47.5, counsel for the United States,
defendant-appellee, states that there were two appeals of this civil action

previously before the Court. The first appeal, a consolidated appeal captioned as *Mid Continent Steel & Wire, Inc. v. United States*, Nos. 2018-1229, 2018-1251, was heard by a panel composed of Circuit Judges Dyk, Linn, and Taranto and decided on October 17, 2019, in *Mid Continent Steel & Wire, Inc. v. United States*, 940 F.3d 662 (Fed. Cir. 2019) (*Mid Continent III*).¹ The second appeal, *Mid Continent Steel & Wire, Inc. v. United States*, No. 2021-1747, was heard by a panel composed of Circuit Judges Newman, Lourie, and Taranto and decided on April 21, 2022, in *Mid Continent Steel & Wire, Inc. v. United States*, 31 F.4th 1367 (Fed. Cir. 2022) (*Mid Continent V*) (Appx965-973).

Counsel is aware of no appeal currently pending in the Court that may be directly affected by the Court's decision in this appeal.

¹ Mid Continent Steel & Wire, Inc. (Mid Continent), plaintiff-appellee, and defendants-appellants, PT Enterprise Inc., Pro-Team Coil Nail Enterprise Inc. (Pro-Team), Unicatch Industrial Co., Ltd., WTA International Co., Ltd., Zon Mon Co., Ltd., Hor Liang Industrial Corporation, President Industrial Inc., and Liang Chyuan Industrial Co., Ltd. (collectively, "PT"), filed separate actions in the U.S. Court of International Trade (trial court) challenging aspects of the U.S. Department of Commerce's (Commerce) final determination in the less-than-fair-value investigation of certain steel nails from Taiwan. *Certain Steel Nails from Taiwan*, 80 Fed. Reg. 28,959 (Dep't of Commerce May 20, 2015) (final determination of sales at less than fair value) (*Final Determination*). The trial court consolidated the actions and, but for the issue before the Court in this appeal, all issues raised in Mid Continent's and PT's complaints were resolved. *See Mid Continent Steel & Wire, Inc. v. United States*, 219 F. Supp. 3d 1326 (Ct. Int'l Trade 2017) (*Mid Continent I*); *Mid Continent Steel & Wire, Inc. v. United States*, 273 F. Supp. 3d 1161 (Ct. Int'l Trade 2017) (*Mid Continent II*); *Mid Continent III*, 940 F.3d 662; *Mid Continent Steel & Wire, Inc. v. United States*, 495 F. Supp. 3d 1298 (Ct. Int'l Trade 2021) (*Mid Continent IV*).

Counsel is also aware of several cases pending in the trial court that will be directly affected by the Court's decision in this appeal: *Resolute FP Canada Inc. v. United States*, Court No. 23-00095 (Ct. Int'l Trade); *Maquilacero S.A. de C.V. et al. v. United States*, Court No. 23-00091 (Ct. Int'l Trade); *Garg Tube Export LLP v. United States*, Court No. 21-00169 (Ct. Int'l Trade); and *Government of Canada, et al. v. United States*, Court No. 23-00187 (Ct. Int'l Trade).

STATEMENT OF THE ISSUE

The sole question before the Court in this appeal is whether, within the context of Commerce's differential pricing analysis, Commerce's use of a simple average in the denominator of the Cohen's *d* coefficient is reasonable, supported by substantial evidence, and in accordance with law. *See Mid Continent V*, 31 F.4th at 1369-70; Applnt. Br. 3. As the trial court correctly held, the answer to this question is yes: Commerce's choice to use a simple average in the denominator of the Cohen's *d* coefficient in the context of the Commerce's differential pricing analysis is reasonable, supported by substantial evidence, and otherwise in accordance with law.

There are *no* other issues before the Court in this appeal and, as discussed below, the attempt by the *Amici Curiae*, the Government of Canada, Canfor Corporation, Canadian Forest Products, Ltd., Canfor Wood Products Marketing, Ltd., J.D. Irving, Limited, Resolute FP Canada Inc., Tolko Industries Ltd., Tolko

Marketing and Sales Ltd., and West Fraser Mills Ltd. (collectively, “Canada”), to raise additional issues and expand the scope of this appeal should be rejected.

STATEMENT OF THE CASE²

This appeal concerns determinations made by Commerce in the less-than-fair-value investigation of certain steel nails from Taiwan. *Certain Steel Nails from Taiwan*, 80 Fed. Reg. 28,959 (Dep’t of Commerce May 20, 2015) (final determination of sales at less than fair value) (*Final Determination*), and the accompanying Issues and Decision Memorandum (Appx384-442, Appx443-446, Appx1071-1121); *Certain Steel Nails From the Republic of Korea, Malaysia, the Sultanate of Oman, Taiwan, and the Socialist Republic of Vietnam*, 80 Fed. Reg. 39,994 (Dep’t of Commerce July 13, 2015) (antidumping order) (Appx447-450). PT appeals the trial court’s judgment in *Mid Continent Steel & Wire, Inc. v. United States*, 680 F. Supp. 3d 1346 (Ct. Int’l Trade 2024) (*Mid Continent VII*) (Appx1-21; Appx22-24), sustaining Commerce’s final redetermination (Appx2401-2467).

² We disagree with PT’s characterizations of certain facts upon which Commerce’s determination and the trial court’s judgment were based. *See* Applnt. Br. 3-19. We also include facts that are necessary to address the arguments raised in GOC’s brief. Therefore, we provide this statement of the case as permitted by Federal Circuit Rule 28(b).

I. Mid Continent V Background

A. Initial Administrative Proceeding

In response to a petition filed by Mid Continent, Commerce initiated a less-than-fair-value investigation of certain steel nails from Taiwan. *Certain Steel Nails from India, the Republic of Korea, Malaysia, the Sultanate of Oman, Taiwan, the Republic of Turkey, and the Socialist Republic of Vietnam*, 79 Fed. Reg. 36,019 (Dep't of Commerce June 25, 2014) (initiation) (Initiation Notice). Commerce selected Taiwanese exporter PT Enterprise and its affiliated producer Pro-Team, among others, as mandatory respondents in the investigation. *See Certain Steel Nails from Taiwan*, 79 Fed. Reg. 78,053, 78,054 (Dep't of Commerce Dec. 29, 2014) (negative prelim. determination of sales at less than fair value, postponement of final determination) (Appx58-59) and accompanying Preliminary Decision Memorandum (Appx75-77).

Commerce issued its final determination in the less-than-fair-value investigation covering certain steel nails from Taiwan, finding that certain steel nails from Taiwan were being sold at less than fair value in the United States. *See Certain Steel Nails From Taiwan*, 80 Fed. Reg. 28,959 (Dep't of Commerce May 20, 2015) (final determination of sales at less than fair value) (Appx84-94) and the accompanying Issues and Decision Memorandum (Appx348-406). As it had done in the preliminary determination, Appx58-59, Appx75-77, Commerce continued to

employ the differential pricing analysis. Appx84-94, Appx348-406. However, while Commerce's determination that there existed a pattern of prices for comparable merchandise that differed significantly among purchasers, regions, or time periods remained unchanged, unlike in the preliminary determination, Commerce determined that such differences could not be accounted for under the average-to-average (A-to-A) comparison method. Appx77; Appx365-367; PT Final Sales Analysis Memorandum (May 13, 2015) (Appx96). Accordingly, based on its finding that 42.27 percent of PT's U.S. sales had passed the Cohen's *d* test, Commerce applied the average-to-transaction (A-to-T) method to PT's U.S. sales that passed the Cohen's *d* test, and it applied the A-to-A method to PT's U.S. sales that did not pass the Cohen's *d* test. Appx365-367. Commerce determined an estimated weighted-average dumping margin of 2.24 percent for PT.³ Appx88.

Following the affirmative injury determination by the U.S. International Trade Commission, Commerce issued an antidumping duty order covering certain steel nails from Taiwan. *Certain Steel Nails From the Republic of Korea, Malaysia, the Sultanate of Oman, Taiwan, and the Socialist Republic of Vietnam*, 80 Fed. Reg. 39,994 (Dep't of Commerce July 13, 2015) (order) (Appx407-417).

³ PT's weighted-average dumping margin decreased to 2.16 percent following the recalculation of its general and administrative expenses on remand; however, the results of Commerce's differential pricing analysis did not change. *Mid Continent II*, 273 F. Supp. 3d at 1164-1165.

B. Judicial Challenges To Commerce's Determinations

PT initially challenged three aspects of Commerce's differential pricing methodology as contrary to law: (1) Commerce's application of the "mixed" comparison method, including what PT referred to as "double zeroing," to calculate PT's estimated weighted-average dumping margin; (2) Commerce's use of the Cohen's *d* test thresholds to determine whether prices differed significantly; and (3) Commerce's use of a simple average to calculate the denominator for the Cohen's *d* coefficient. The trial court sustained Commerce's final determination with respect to the challenged differential pricing issues. *Mid Continent I*, 219 F. Supp. 3d at 1344, 1351; *Mid Continent II*, 273 F. Supp. 3d at 1163.

The Court affirmed the trial court's judgment in all respects, except one – it vacated the trial court's judgment affirming Commerce's determination to use a simple average to calculate the denominator of the Cohen's *d* coefficient and remanded that issue to secure further explanation from Commerce. *Mid Continent III*, 940 F.3d at 665, 675.

Following the remand order, Commerce issued a draft for the second remand redetermination, Appx460-476, and reopened the record for the submission of new factual information, which included sections of relevant academic literature. See Appx448-521 (Paul D. Ellis, *The Essential Guide to Effect Sizes* (2010) (Ellis); Appx523-561 (Jacob Cohen, *Statistical Power Analysis for the Behavioral*

Sciences (2d ed. 1988) (Cohen)); Appx563-573 (Robert Coe, *It's The Effect Size, Stupid: What Effect Size Is And Why It Is Important*, Paper presented at the Annual Conf. of British Educational Research Ass'n (Sept. 2002) (Coe)). After PT and Mid Continent submitted additional factual information, Appx580-735, Appx740-805, Commerce released the second remand redetermination, explaining the bases for its continued use of a simple average to calculate the denominator of the Cohen's *d* coefficient. Appx806-854.

The trial court sustained Commerce's second remand redetermination. *Mid Continent IV*, 495 F. Supp. 3d 1298. On appeal, however, the Court vacated *Mid Continent IV*, holding that Commerce had not provided a "reasonable justification for departing from what the acknowledge literature teaches about Cohen's *d*." *Mid Continent V*, 31 F. 4th at 1381 (Appx974-985). The Court concluded that: (1) Commerce had departed from the methodology described in the academic literature when it adopted simple averaging for the denominator of the Cohen's *d* coefficient; (2) Commerce's reasons for the departure, that the pricing behavior of the test and comparison groups was equally genuine and equally rational, had no connection to the purpose of the denominator, *id.*; and (3) nothing in the section of the academic literature from which Commerce's formula is borrowed applies simple averaging to pool standard deviation estimates for different-sized groups, *id.* at 1372, 1380. Based on these conclusions, the Court again remanded,

instructing Commerce to “either provide an adequate explanation for its choice of simple averaging or make a different choice, such as use of weighted averaging or use of the standard deviation for the entire population.” *Id.* at 1381.

II. Post-Mid Continent V Proceedings

Following the third remand redetermination, Commerce again explained how the academic literature supports the use of a simple average to calculate the denominator of the Cohen’s d test. Appx986-1007. The trial court, however, found that Commerce’s determination required further explanation and remanded the action. *Mid Continent Steel & Wire, Inc. v. United States*, 628 F. Supp. 3d 1316-18, 1324 (Ct. Int’l Trade 2023) (*Mid Continent VI*) (Appx1719-1728).

On the fourth remand, Commerce issued a draft redetermination and invited and received comments from Mid Continent and PT. *See* Appx1761-2400; Appx1729-1747. Thereafter, Commerce issued the fourth redetermination, explaining the underlying basis for its use of a simple average of the standard deviation of full populations to calculate the denominator of the Cohen’s d coefficient. Appx2401-2467.

Commerce acknowledged that in the academic literature, Cohen’s equation 2.3.2, calls for the use of a simple average to calculate the denominator of the Cohen’s d coefficient when population variances are not equal. Commerce also acknowledged the Court’s understanding that equation 2.3.2 applies to sample

sizes that are equal. Commerce explained that use of the simple average can be reasonably explained when the sample sizes are equal because the “reliability” of the estimated standard deviations of the sampled groups is equal. This is so because the size of a sample is an indicator of reliability. Therefore, when sample sizes are equal, the averaged estimated standard deviations of the samples are equally reliable.

Commerce’s Cohen’s d test is based on Cohen’s equation 2.3.2. As applied by Commerce in its differential pricing analysis, the Cohen’s d test involves comparing the prices of the “test group” of a respondent’s sale prices to the “comparison group” of the respondent’s other prices of comparable merchandise by region, purchaser, and time period. Although the test and comparison groups are not of equal sizes, Commerce explained that each group represents a full, complete population, *i.e.*, all the U.S. prices to that group. Consequently, any parameter calculated based on the population, e.g., the standard deviation, of each group is 100 percent reliable in representing the actual value of that parameter for the population. In other words, the calculated value of the parameter is not an estimate of that value but is the actual value of the parameter. Therefore, based upon the general statistical principle of reliability gleaned from the academic literature, Commerce determined that the use of a simple average to calculate the denominator of the Cohen’s d coefficient within the context of Commerce’s

Cohen's d test is reasonable because the calculated standard deviation of the test group and comparison group is equally reliable. Appx2410-2413. Commerce also explained why the use of a single standard deviation of all prices in both the test group and the comparison group is not appropriate for use as the denominator of Commerce's Cohen's d coefficient. Appx2413-2417. In the end, Commerce concluded that the use of a simple average is appropriate. Appx2421-2463.

The trial court sustained Commerce's fourth remand redetermination, holding that Commerce's explanation for departing from what the academic literature teaches about the Cohen's d coefficient is reasonable. *Mid Continent VII*, 680 F. Supp. 3d 1346 (Appx1-21; Appx22-24). The trial court found that Commerce had identified where simple averaging is supported by the literature, extrapolated a rationale for why simple averaging is appropriate for the denominator in Commerce's Cohen's d test, and then applied that rationale to the circumstances before Commerce. *Id.* at 1353. The trial court further found that the principle that Commerce derives from the academic literature leads to a logical conclusion that simple averaging in Commerce's Cohen's d test is a reasonable choice. *Id.* Finally, the trial court explained that, although there may be other reasonable alternatives, it could find no fault with Commerce's logic. *Id.* This appeal followed.

SUMMARY OF THE ARGUMENT

The trial court's judgment should be sustained because Commerce's use of a simple average for denominator of the Cohen's d coefficient is reasonable and supported by substantial evidence.

Because the academic literature does not speak directly to Commerce's Cohen's d test, Commerce used the general principle of reliability presented in the academic literature to determine that simple average is appropriate for the denominator of Commerce's Cohen's d coefficient. Dr. Cohen's equation 2.3.2. calls for a simple average when populations' standard deviations are not equal, *i.e.*, $\sigma_A \neq \sigma_B$. With the understanding that equation 2.3.2 applies when sample sizes are equal, *i.e.*, $n_A = n_B$, a simple average is reasonable because sample size is an indicator of reliability, and, thus, the averaged estimated standard deviations are equally reliable. Extending this to full populations, as with the test and comparison groups in Commerce's Cohen's d test, the actual standard deviations are also equally reliable, and, thus, Commerce finds that a simple average is appropriate for its Cohen's d test.

Next, Commerce addressed the Court's suggested use of a single standard deviation commingling the data from both the test and comparison groups. Commerce explained that Dr. Cohen's original formulation defines the denominator as the common standard deviation as he assumes that the standard

deviations of each population, A and B, are equal. The single standard deviation of commingled populations A and B is not the same as the standard deviation of population A or the standard deviation of population B. Commerce explained that the single standard deviation of commingled groups not only includes the variance in the data within each group, but also the difference between the groups, and, thus, single standard deviation will vary with the difference in the means between the two groups. Thus, the single standard deviation is not an appropriate measure to “contextualize” the difference in the means. Moreover, to carry out the statute and determine whether U.S. sales prices differ significantly between purchasers, regions, or time periods, Commerce views the test and comparison groups, which are separate populations with unequal variances, as distinct and should not be commingled.

PT’s arguments are unavailing. PT’s visual examples show how weight averaging works but do not express whether the data represented in the graphics differ significantly or demonstrate that Commerce’s approach is unreasonable. Commerce is not calculating dumping margins using the Cohen’s d test, but rather determining whether U.S. prices differ significantly. PT makes multiple assertions about reliability that are not explained and are otherwise not supported. PT also invokes ANOVA, which is relevant to Dr. Cohen’s f coefficient, a distinct concept from Dr. Cohen’s d coefficient.

Finally, Canada presents new issues that are beyond the scope of this appeal, makes new arguments that were not raised before Commerce, and it relies on academic literature that is not on the record. Simply put, Canada cannot expand the scope of PT's appeal by raising new issues and arguments based upon factual information that is not on the record of this proceeding, and Canada should not be permitted to circumvent the principles of exhaustion and waiver. Consequently, the Court should disregard Canada's new issues and arguments, and non-record citations.

ARGUMENT

I. Standard Of Review

The Court gives “due respect to the informed opinion of the Court of International Trade.” *Novosteel SA v. United States*, 284 F.3d 1261, 1269 (Fed. Cir. 2002) (internal quotation marks omitted). However, to determine whether the trial court correctly applied the standard of review in reaching its decision, the Court “must apply anew the statute’s express standard of review to the agency’s determination.” *Fujitsu Gen. Ltd. v. United States*, 88 F.3d 1034, 1038 (Fed. Cir. 1996) (citations omitted). The statute provides that the Court must sustain “any determination, finding or conclusion” found by Commerce unless it is “unsupported by substantial evidence on the record, or otherwise not in accordance with law.” 19 U.S.C. § 1516a(b)(1)(B)(i). PT’s suggestion that the Court apply a

lesser standard to Commerce’s fourth remand determination, Applnt. Br. 25, should be disregarded as contrary to the statutory standard of review that applies to “any” determination, which includes those determinations following a court’s remand order to which the agency’s decision making must comport. *See SEC v. Chenery Corp.*, 332 U.S. 194, 199-201 (1947).

The Court has determined that the appropriate standard for reviewing Commerce’s differential pricing analysis and its components is reasonableness. *See Stupp Corp. v. United States*, 5 F.4th 1341, 1353 (Fed. Cir. 2021) (“Our precedents make clear that the relevant standard for reviewing Commerce’s selection of statistical tests and numerical cutoffs is reasonableness, not substantial evidence”) (citations omitted); *Mid Continent III*, 940 F.3d at 667 (“In carrying out its statutorily assigned tasks, Commerce has discretion to make reasonable choices within statutory constraints.” (collecting cases)); *Borusan Mannesmann Boru Sanayi Ve Ticaret A.S. v. Am. Cast Iron Pipe Co.*, 5 F.4th 1367, 1374-75 (Fed. Cir. 2021) (recognizing “deference {that} is both greater than and distinct from that accorded the agency in interpreting the statutes it administers, because it is based on Commerce’s technical expertise”); *Smith-Corona Group v. United States*, 713 F.2d 1568, 1582 (Fed. Cir. 1983) (holding that “review of the statute reveals tremendous deference to the expertise of {} Commerce in administering the antidumping law”).

II. Commerce's Determination To Use A Simple Average For The Denominator Of The Cohen's d Test Is Reasonable

The trial court's judgment that Commerce's choice to use a simple average to calculate the denominator of the Cohen's d coefficient in its differential pricing analysis is reasonable and should be affirmed.

A. Commerce's Differential Pricing Analysis And Cohen's d Test

After a determination has been made that imported merchandise is being sold in the United States at less than fair value, Commerce compares normal value to the export price (or constructed export price) of subject merchandise to calculate a weighted-average dumping margin for individually examined exporters and producers. 19 U.S.C. § 1677(35)(A); *Apex Frozen Foods Private Ltd. v. United States*, 144 F. Supp. 3d 1308, 1315 (Ct. Int'l Trade 2016), *aff'd*, 862 F.3d 1337, 1344 (Fed. Cir. 2017). In accordance with the statute, Commerce ordinarily uses either the A-to-A or the transaction-to-transaction (T-to-T) methods to calculate the estimated weighted-average dumping margin in an investigation. When (1) there exists a pattern of prices that differ significantly for comparable merchandise among purchasers, regions, or time periods and (2) the A-to-A or T-to-T comparison methods cannot account for such differences, Commerce may use the average-to-transaction (A-to-T) method to calculate the estimated weighted-average dumping margin as an alternative comparison method. 19 U.S.C. § 1677f-1(d)(1)(A) -(B); *Apex*, 144 F. Supp. 3d at 1315. With the A-to-T method,

Commerce compares weighted-average normal values with the export price or constructed export price of individual sales.

The statute does not provide a Congressionally mandated procedure or methodology to determine whether (1) there exists a pattern of prices that differ significantly for comparable merchandise among purchasers, regions, or time periods (the pattern requirement) or (2) one of the standard comparison methods cannot account for such differences (the meaningful difference requirement). Therefore, Commerce conducts the differential pricing analysis to implement the dictates of the statute. Appx2405.

Commerce's differential pricing analysis contains three tests – the Cohen's *d* test, the ratio test, and the meaningful difference test. The ratio test and the meaningful difference test are not at issue in this appeal. Following the direction of 19 U.S.C. § 1677f-1(d)(1)(B)(i), Commerce's Cohen's *d* test examines whether the U.S. prices to a given purchaser, region, or period of time differ significantly from the U.S. prices of comparable merchandise to other purchasers, regions, or time periods. Appx2405, 2455.

For each category *{i.e., purchaser, region, or period of time}*, Commerce segregates sales into subsets, with one subset becoming the test group, and the remaining subsets being combined as the comparison group. {See Differential Pricing Analysis; Request for Comments, 79 Fed. Reg. 26,720, 26,722 (Dep't of Commerce May 9, 2014)}. Commerce then calculates the means and standard deviations of the test and comparison groups.

Id. Commerce finally calculates a d coefficient {,} {the “Cohen’s d coefficient,”} by dividing the difference in the groups’ means by the square root of the average of the squared standard deviations of each group. { } {See Appx2406} (citing Cohen at 20). Commerce finds the average of the squared standard deviations by adding them together and dividing by two, referring to the result as a “simple average.” See id. Commerce does not account for the differences in the size of each group, i.e., use a “weighted average.” {Appx2406}.

Mid Continent VII, 680 F. Supp. at 1351 (footnote omitted). Thereafter, Commerce compares the prices in each test group with the prices in the comparison group for each category (*i.e.*, purchaser, region, or period of time) and assigns a d coefficient. *Id.*

Commerce’s Cohen’s d coefficient is used to evaluate the extent to which the weighted-average net prices (*i.e.*, the mean of the net prices) to a particular purchaser, region, or period of time (*i.e.*, the test group) differ from the weighted-average net prices of all other sales of comparable merchandise (*i.e.*, the comparison group). Appx76. The Cohen’s d coefficient is calculated as a ratio, where the numerator is the difference in the means of the net prices and the denominator is the “standard deviation.” Appx76; *see also* Appx534. In this way, the difference in the mean net prices between the test group and the comparison group is measured relative to the variation in the prices within both groups. Commerce calculates the denominator of the effect size as a “simple average,” *i.e.*,

the “root mean square,” of the standard deviations of the test and comparison group. Appx2406; Appx545-546.

Commerce’s Cohen’s d test is based on a measure of “effect size,” which is the measure of the practical significance of the difference in two means, and it is a concept expounded by Dr. Cohen in his textbook on statistical power analysis.

Appx2405-2406. *See also* Appx567 (“Effect size quantifies the size of the difference between two groups {} and may therefore be said to be a true measure of the significance of the difference.”). The effect size, the “Cohen’s d coefficient,” is the ratio of the difference in the means of two groups, divided by the “standard deviation,” *i.e.*, the variance in the underlying data. Appx2406.

The significance of the difference in the means is determined by one of three fixed thresholds proposed by Dr. Cohen: small, medium, or large. Appx530 Of these thresholds, the large threshold, 0.8, provides the strongest indication that there is a significant, practical difference between the means of the net prices of the test and comparison groups. Appx76. Dr. Cohen’s thresholds have been widely adopted. *Mid Continent III*, 940 F.3d at 673. *See also* Appx530 (“Although arbitrary, the proposed conventions will be found to be reasonable by reasonable people.”). If the d value of the test group is equal to or greater than the “large” threshold, 0.8, the observations within the test group are said to have “passed” the Cohen’s d test. Together, these steps allow Commerce to determine whether U.S.

prices differ significantly among purchasers, regions, or periods of time. 19 U.S.C. § 1677f-1(d)(1)(B)(i).

B. The Principle That Commerce Derived From The Academic Literature Leads To A Logical Conclusion That Use Of A Simple Average To Calculate The Denominator Of Commerce's Cohen's *d* Coefficient Is Reasonable

The academic literature does not speak directly to Commerce's Cohen's *d* test that, as applied, involves comparing the prices of the "test group" of a respondent's sale prices to the "comparison group" of the respondent's sale prices by region, purchaser, and time period. The remand redetermination contains a comprehensive and reasonable explanation for Commerce's determination to use a simple average to calculate the denominator of the Cohen's *d* coefficient within the context of Commerce's Cohen's *d* test. Commerce's remand determination also provides a reasonable justification for departing from what the academic literature on the record of this proceeding instructs about the use of Cohen's *d* formula when sample sizes differ. Therefore, as the trial court did, this Court should sustain Commerce's determination to use a simple average in its Cohen's *d* test.

1. The Standard Deviations Of Two Different Sized Full Populations Are Equally Reliable Just As The Standard Deviations Of Two Samples Of Equal Size Are Equally Reliable

The Court determined in *Mid Continent V* that the academic literature on effect size uniformly instructs that use of the "pooled standard deviation" estimate

involves weight averaging, and the academic literature does not apply simple averaging to a pooled standard deviation estimate for different-sized sampled groups. 31 F. 4th at 1370, 1377-78 (“a situation in which the two groups at issue are of the same size. (‘CASE 2: $\sigma_A \neq \sigma_B$, $n_A = n_B$ ’).”) (cleaned up). Rather, the Court stated, the academic literature “indicates that the ideal denominator is the full population’s standard deviation,” but Commerce “did not use the standard deviation of all the data for its denominator” and instead used a simple average of the standard deviations for the two groups. *Id.* at 1378. The Court held that, by adopting simple averaging for the denominator of the Cohen’s *d* coefficient, Commerce had departed from the methodology described in the academic literature, which requires $n_A = n_B$, *i.e.*, sampled groups of equal size. *Id.*

The Court also criticized Commerce’s stated reasons for the departure from the methodology described in the academic literature, that the pricing behavior of each group was equally genuine and equally rational, stating that Commerce’s reasons had no connection to the purpose of the denominator (which is to provide a dispersion figure for the more general pool that served as yardstick for deciding on the significance of difference in mean prices of two groups). *Id.* at 1379. The Court determined that Dr. Cohen’s equation 2.3.2, which Commerce uses, applies to groups that have the same sample size ($n_A = n_B$) and different standard deviations ($\sigma_A \neq \sigma_B$). *Id.* at 1372, 1378 (stating that the groups referenced are “sampled

groups”). However, Commerce used Dr. Cohen’s equation 2.3.2 without group-size weighting even when, unlike the situation described in the section in Cohen from which Commerce’s formula is borrowed, the groups are of different sizes. *Id.* at 1372. The Court concluded that nothing in the section of the academic literature from which Commerce’s formula is borrowed applies simple averaging to pool standard deviation estimates for different-sized “groups.” *Id.*

In the fourth remand redetermination, Commerce explained that the use of a simple average where sample sizes are equal stems from the equal reliability of standard deviations in samples of equal sizes. Appx2413. Commerce’s focus on reliability derives from Dr. Cohen’s presentation of the parameters of the statistical power analysis and description of the “reliability of sample results and sample size” and the underlying general statistical principal that the reliability of sample results is predicated on sample size. Appx2410 (quoting Appx527 (“{Reliability} is *always* dependent upon the size of the sample.”) (emphasis in original)). As a part of Dr. Cohen’s power analysis, the sample size is a gauge of the reliability of sample results. *Id.* Commerce concluded that “the larger the sample size, the more reliable the sample results.” Appx2410. Against this backdrop, Commerce addressed the Court’s conclusion in *Mid Continent V* that Dr. Cohen’s equation 2.3.2 calls for the use of a simple average when populations variances are not

equal, but the samples sizes are equal, *i.e.*, $\sigma_A \neq \sigma_B$, $n_A = n_B$. Appx2411 (citing *Mid Continent V*, 31 F. 4th at 1378).

The decision in *Mid Continent V* reflects the Court’s understanding that the use of the simple average of differing standard deviations, Dr. Cohen’s equation 2.3.2, applies to an analysis involving sampled data because it is part of a power analysis that involves sampled data, including the *t*-test to evaluate the statistical significance of the results. Appx2411 (citing Appx545-546). In considering the Court’s understanding of equation 2.3.2, Commerce explained that, because the sample sizes are equal, the “reliability” of the estimated standard deviation for each of the sampled groups is also equal. Appx2410-2411. As a result, the simple average is appropriate when there is equality in sample size and reliability. Appx2411. “In other words, when the sample sizes of the two groups are equal, then the reliability of the estimates of the standard deviations are the same, and it is appropriate to give equal weights, *i.e.*, a simple average, when averaging the two standard deviations to calculate the denominator of the Cohen’s *d* coefficient.” *Id.*

The Court also stated its understanding that “when the sampled groups have unequal sizes $\{i.e., n_A \neq n_B\}$, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging.” *Mid Continent V*, 31 F.4th at 1378 (referencing Appx557; Appx503-504; Appx568). Regarding the use of a weighted average of unequal sample sizes, Commerce explained that “the

standard deviation of the group with the larger sample size (*i.e.*, sales volume) is given more weight than the group with the smaller sample size.” Appx2411.

When the sample size of one group is larger than the other group, the “reliability” of the estimated standard deviation for that group will also be greater. *Id.* Because the estimated standard deviation of the group with the larger sample size has greater reliability, more weight is given to that estimated standard deviation, and less weight is given the estimated standard deviation for the group with the smaller sample size, where the weights reflect the relative reliability of the estimated standard deviations from the two groups. Appx2412.

Commerce determined that, because it uses the full populations of U.S. sale prices for both the test and the comparison groups in its Cohen’s *d* test, the standard deviations calculated for these groups have a reliability of 100 percent, *i.e.*, “the closeness with which {the calculated value} can be expected to approximate the relevant population value.” Appx2412 (quoting Appx527) (cleaned up). The calculated standard deviations are equally reliable, and because the “use of the simple average when the sample sizes are equal reflects that the calculated parameters used to calculate the Cohen’s *d* coefficient are equally reliable,” Commerce determined that it is reasonable to combine the standard deviations for Commerce’s Cohen’s *d* test using a simple average. Appx2413.

2. PT's And Canada's Challenges To Commerce's Reliability Analysis As Supporting The Use Of A Simple Average Lack Merit

PT and Canada argue that Commerce's use of a simple average is not reasonable and is contrary to the academic literature. Applnt. Br. 26-28; Amici Br. 18-23. They complain that Commerce cannot use the academic literature to create its Cohen's *d* test and then ignore or depart from the literature's mandate that the standard deviation cannot be based on the simple average of two groups of unequal size. PT Br. 19-20, 22, 26-28, 49-54; Amici Br. 19-22. They also argue that Commerce's "reliability" rationale is not reasonable. Applnt. Br. 49-54; Amici Br. 23-27. Contrary to PT's and Canada's arguments, Commerce has provided a fulsome and reasonable explanation for its choice to use a simple average based on the general principle of reliability.

As the Court explained, "Commerce's job is not to follow a statistical test as explained in published literature for its own sake, but to implement the statutory mandate to determine when prices of certain groups 'differ significantly.'" *Mid Continent V*, 31 F.4th 1380-81 (quoting 19 U.S.C. § 1677f-1(d)(1)(B)(i)). Indeed, "{i}n implementing a statutory mandate, an agency is not duty-bound to follow published literature. . . ." *Id.* at 1381. Thus, the Court explicitly acknowledged that Commerce is not limited to the academic literature in supporting its determination, stating that "Commerce needs a reasonable justification for

departing from what the acknowledged literature teaches about Cohen's d ." *Id.*

As the trial court correctly noted, Commerce explained that it "is not relying solely upon the academic literature to support its choice {of a simple average}, but rather argues that the principle it derives from the academic literature leads to a logical conclusion that simple averaging in this case is a reasonable choice." *Mid Continent VII*, 680 F. Supp. 3d at 1353 (citing Appx2412-2413, Appx2422-2425). Commerce provided an explanation that identifies where simple averaging is supported by the literature (equation 2.3.2), extrapolates a rationale for why simple averaging is appropriate (reliability of equal sample sizes), and then extends that equal reliability rationale to the full populations, *i.e.*, the test and comparison groups, in the Cohen's d test.

PT argues that Commerce has not justified that a simple average, rather than a weighted average, should be used, and that equal reliability does not mean that each standard deviation should be accorded equal weight. Applnt. Br. 22, 49-54. This argument is unavailing because Commerce explained why the use of a simple average, rather than a weighted average, is reasonable for Commerce's Cohen's d test. Appx2410-2413. Commerce explained that "{w}ith the weighted average, the standard deviation of the group with the larger sample size (*i.e.*, sales volume) is given more weight than the group with the smaller sample size" and that "{i}f the sample size of group A is larger than the sample size of group B, then the

reliability of the standard deviation of group A will be greater than the reliability of group B.” Appx2411. Consequently, the standard deviation of group A has more reliability and is given more weight than the standard deviation of group B.

Appx2412. And “{b}ecause the group with the larger sample size has greater reliability, the weights reflect the relative reliability of the standard deviations from the two groups.” Appx2412. Conversely, if the sample sizes of the two groups are equal, then the reliability of the estimated standard deviations are also the same, and it is appropriate to give equal weights, *i.e.*, a simple average, when averaging the two standard deviations. Appx2411. In Commerce’s application, the reliability of the calculated standard deviations, based on the full population of U.S. prices to each group, are each 100 percent, are identical, and “Commerce finds that it is reasonable to weight these standard deviations equally, *i.e.*, a simple average, as presented in Dr. Cohen’s equation 2.3.2, just as when the reliability is equal for standard deviations based on sampled data with equal sample sizes.” Appx2412. Commerce’s explanation is logical, and it demonstrates that a simple average is reasonable when the reliability of the standard deviations for the two groups are equal. Therefore, PT’s argument fails.

Suggesting that the reliability of the groups stays the same regardless of which standard deviation yardstick is used, PT argues that the Commerce has not provided an adequate explanation for its use of a simple average over a weighted

average or a standard deviation of the entire population. Applnt. Br. 49. This argument is unavailing because Commerce must only demonstrate that a simple average is reasonable, not that it is the only reasonable option, or even that it is the most reasonable option. *See Borusan*, 5 F.4th at 1374 (explaining reasonable minds may differ on the outcome, but that does not mean a determination fails or is unreasonable).

Canada asserts that Dr. Cohen weights by size, not reliability, and reliability is only relevant in the context of sampled data and not with full populations. Amici Br. 24, 26. Canada further argues that Commerce merely creates a “truism” that the calculated “observed” parameter is always equal to the “true” parameter when Commerce finds that it is using the full population of data. *Id.* at 26. However, reliability is relevant. Indeed, a weighted average is specifically weighted by sample size. Appx487 (“Another approach, which is recommended if the groups are dissimilar in size, is to weight each group’s standard deviation by its sample size.”). Sample size is an indicator of reliability, and when Dr. Cohen defines the four parameters of statistical inference, “sample size(n)” represents the reliability of the sample results. Appx2410 (citing Appx527, Appx531). Interestingly, citing Cohen at 6 (Appx1925), Canada accepts that reliability describes the difference between an “observed” parameter and the “true” parameter. Amici Br. at 25. Even so, while accepting the concept of reliability,

Canada fails to explain how the concept does not also apply to populations. In other words, Canada does not explain how a calculated “observed” parameter based on all the data does not always equal the “true” parameter, *i.e.*, which is 100 percent reliable. Dr. Cohen states that reliability “is the closeness with which {a value} can be expected to approximate the relevant population value.” Appx527. By extension then, reliability is 100 percent when the relevant populations’ values (*i.e.*, the test group and comparison group) are not approximations or estimations.

Even though PT agrees that a simple average is appropriate with equal sample size because this is the consequence of weight averaging with equal weights, PT argues that equal reliability does not mean equal weighting. Applnt. Br. 49-50. Although the academic literature does not directly explain the reason for using a weighted average when the standard deviations are based on sampled data, Dr. Cohen explains that standard error is inversely proportional to the sample size and “one conventional means for assessing the reliability of a statistic is the standard error (SE) of the statistic.” Appx527. Further, Dr. Cohen confirms that “{t}he nature of the dependence of reliability upon **n** {*i.e.*, the sample size) is obvious from the illustrative formulas, and, indeed, intuitively. The larger the sample size, other things being equal, the smaller the error and the greater the reliability or precision of the results.” *Id.* From this, Commerce reasonably inferred that the weighting of the estimated standard deviations by sample sizes

can logically be explained by the relative reliability of the estimated standard deviations.

PT also contends that just because a simple average can be used when the sample sizes are equal does not mean that when full populations have unequal sizes that a simple average can be used. Applnt. Br. 51. PT points to no record evidence to support this contention. However, as discussed above, Dr. Cohen draws a direct relationship between sample size and reliability, and Commerce has thoroughly explained how the concept of reliability supports its use of a simple average. Appx2421-2422.

PT states that Dr. Cohen, and the academic literature in general, examined phenomena that were based on “the number of counts in each group,” usually based on persons, each of which constituted a single unit or observation. Applnt. Br. 50-51. Therefore, PT argues that the correct weight for Commerce’s dumping analysis is the quantity sold and not the number of observations, and, thus, Commerce reliance on equal sample sizes “does not support its conclusion.” *Id.* at 51-52. It is true that Commerce generally weights using quantity sold. The academic literature speaks of weighting based on sample size, *e.g.*, the number of sales. *See e.g.*, Appx504. “Sample size” could be the number of sales in the sample, or it could be the quantity sold in the sample, where, given PT’s argument, each kilogram sold would effectively be an “observation.” However, PT’s

argument is irrelevant to the issue before the Court. In *Mid Continent V*, the Court noted that “Commerce rejected weighted averaging altogether, so we do not have before us for review a choice of one basis of weighting rather than another.” 31 F.4th at 1381 n.6. The choice of one basis for weight averaging over another is not before the Court in this appeal either because Commerce rejected weighted averaging altogether. Appx2425 (“Commerce, in this redetermination, also does not adopt one specific basis for weighting as we continue to find that use of a simple average is reasonable.”).

PT and Canada argue that if Cohen, Ellis, and Coe had based their acceptance of a simple average on reliability concerns, they would have so stated. Applnt. Br. 51; Amici Br. 23-24. This argument is unavailing. Dr. Cohen explained that *all* statistical tests depend on the general concept of the reliability of the calculated results based on sampled data, which “*always* depends on the size of the sample.” Appx527. The larger the sample size, the smaller the standard, or sampling, error and the greater the reliability (or precision) of the results. Appx527. Therefore, it was reasonable for Commerce to rely on the concept of reliability to explain why it is appropriate to use simple averaging when sample sizes are equal. Commerce further explained that the reliability of population standard deviations is also equal, and, thus, it is reasonable to also use a simple average. Commerce’s explanation reasonably relies on the general principle of

reliability presented in the academic literature on the record to explain when the use of simple average is warranted, ultimately, “Commerce has determined, consistent with the Court’s prior decisions, that it must go beyond the academic literature to explain our use of a simple average to calculate the denominator of the Cohen’s *d* coefficient.” Appx2461. Accordingly, Commerce properly examined whether reliability supports using simple average when standard deviations are calculated for populations (as opposed to samples).

PT contends that because a sample’s reliability depends on multiple factors, the reliability of samples cannot be readily compared to the reliability of a full population and cannot “be a reason why the standard deviation yardstick of a full population can be based on simple averaging of the SDs of two unequal groups.” Applnt. Br. 52-53. However, PT fails to explain how the definition of reliability changes depending upon whether the data are sampled or constitute the full population. As Dr. Cohen explained, reliability is the closeness that the estimated value based on sampled data can be expected to approximate the actual value of the parameter for the full population. Appx527. The concept of reliability does not change when the value is calculated using all data within the population and the approximation, the potential for error, is eliminated. Reliability simply reaches its logical conclusion when there is no possible error in the calculated value and the reliability of the calculated value is 100 percent.

Furthermore, “{t}he reliability of an estimated parameter based on sampled data may be dependent on many factors, but it is *always* dependent on the number of observations in the sampled data, i.e., the sample size (n).” Appx2421 (emphasis added). *See also* Appx527. Of the four characteristics which contribute to statistical inference, only sample size is determinative of the reliability of the results. Appx527, Appx531. Consequently, the standard deviations calculated on two samples with equal sample size, when compared to each other, are equally reliable. Appx2411. Furthermore, Commerce does not state that the reliability of the estimated standard deviations based on sampled data are “100% reliable,” but rather that these estimated values, *i.e.*, the standard deviations, are equally reliable in relation to each other. Indeed, by definition, the values of samples are not 100 percent reliable because they are “estimates” of the actual values of the standard deviations of the two groups. Appx2411.

Canada similarly asserts that reliability is applied only to sample data when making inferences from the population. *Amici Br.* 25-26. However, because the academic literature on the record assumes sampling, Commerce analyzed the characteristic of sampled data in relation to populations, which is presented in the academic literature on the record. Appx2423 (citing *Mid Continent VI*, 628 F. Supp. 3d at 1325).

PT further argues that the fact that the test and comparison groups each has

zero errors is not related to the size of each group. Applnt. Br. 53. This argument supports Commerce's explanation that the reliability of the standard deviations calculated for populations is not dependent on how many observations are in each population. "Zero errors" means 100 percent reliability. Appx2430. PT fails to explain how "zero errors" differs from "perfectly reliable" or "100% reliable."

When the calculated standard deviations are based on the full populations of prices in each group, irrespective of the number of observations in each group, there are "zero errors," the calculated values are the actual standard deviations of each population, and the ability of the values to represent the standard deviations of each population is perfectly reliable. As Commerce stated in the remand redetermination, "{t}he number of observations in the population does not affect the reliability of a value of a parameter (e.g., the standard deviation) of the population." Appx2423. Thus, as in the case with equal sample sizes, it is reasonable to use a simple average because the standard deviations are "equally reliable" when based on full populations.

Finally, PT attempts to equate Commerce's position regarding equal reliability to Commerce's discussion in the second redetermination of "equally rational" and "equally genuine," which the Court held do not support Commerce's use of a simple average. Applnt. Br. 53-54. PT argues that no distinction exists between reliability, rationality, and genuineness and, therefore, Commerce's

position regarding reliability in this redetermination is not supported. *Id.* at 54.

However, Commerce's arguments about reliability are "based on the characteristic of the data rather than the pricing behavior of the respondents." Appx2427.

Although the data may reveal pricing behavior, reliability is a general statistical concept regarding the data, whereas rationality and genuineness are not. In short, the only commonality between the terms is Commerce's use of the adjective "equally." and, thus PT's argument equating the "rationality" and "genuineness" of a respondent's pricing behavior with the "reliability" of parameters calculated for a group of data lacks merit.

**C. Commerce Reasonably Determined That Use Of A Single
Standard Deviation Is Not Appropriate**

In *Mid Continent V*, the Court suggested that the preferred way to establish the denominator was to "use the standard deviation for the entire population." 31 F. 4th at 1377. Commerce addressed the Court's suggestion that the "seemingly preferred way" to calculate the denominator of the Cohen's *d* coefficient is the standard deviation of the combined prices of both the test and comparison groups. Appx2413-2417.

Commerce explained that Dr. Cohen distinguished between the two populations whose differences were being examined, stating that the denominator of the Cohen's *d* coefficient should be the standard deviation of one of the populations when the standard deviations of the two populations are equal.

Appx2414 (citing Appx534, Appx537). Further, Professor Coe warned that a single standard deviation of the combined data would create problems because the “yardstick” by which the difference in the means is measured would include not only the variances in the data within each of the two groups, but also would reflect the difference between the two groups. Appx2415-2416 (citing Appx568). As a result, the single standard deviation will increase when the difference in the means between the two groups increases despite there being no change in the variances within the two groups. Appx2417. By contrast, both the simple average and weighted average of the standard deviations of the two groups will remain constant even as the difference in the means between the two groups increases or decreases. Appx2417.

PT contends that Commerce ignores that the Court has held that the literature supports the single standard deviation, that the test and comparison groups are not different datasets, and that the differences are arbitrary. Applnt. Br. 54-56. However, far from ignoring the Court’s holding, Commerce explained why a single standard deviation is not appropriate for Commerce’s Cohen’s *d* test. In particular, the academic literature supports that the two groups in Commerce’s analysis are distinct and separate from one another. Appx2414-2415. For example, Dr. Cohen’s formulation of effect size defines the denominator as “either the standard deviation of either population A or the standard deviation of

population B when the standard deviation of population A is assumed to be equal to the standard deviation of population B.” Appx2414 (citing Appx534, Appx537). Therefore, the denominator in Dr. Cohen’s equation 2.2.1 and 2.2.2 is either the standard deviation of population A or the standard deviation of population B, but it is not the standard deviation of populations A and B combined. Appx2414. Moreover, when Professor Coe discusses standard deviation, he first suggests using the standard deviation of one group, Glass’s delta or Δ , which is the standard deviation of the control group (here, the comparison group). Appx568. *See also* Appx487 (explaining that Glass’s delta is the standard deviation of the control group). Professor Coe then states that the pooled estimate of standard deviation is “an average of the standard deviations of the experimental and control groups (Equation 4).” Appx568. Professor Coe noted “that this is not the same as the standard deviation of all the values in both groups ‘pooled’ together.” Appx568.

Commerce further explained that “whereas the pooled standard deviation reflects only the variation in the data within each group, the ‘single standard deviation’ not only reflects the variation of the data within each group, but also the difference in the means between the two groups.” Appx2417. This means that even when there is no change in the variance or dispersion in the data within each of the two groups, an increase in the difference of the means would increase the

single standard deviation. Appx2417. The value of a pooled standard deviation, whether simple or weighted average, will remain constant because it is based upon the relationship or spread of data within each group, whereas the value of a single standard deviation will vary with the difference in the means between the two groups. Appx2417, Appx2441-2443, Appx2465-2467 (showing that when the difference in the means between the two groups vary but the variance of each group remains the same, the weighted average and simple average in blue and purple respectively remain constant while the full population single standard deviation in brown varies). “Therefore, the option to use a single standard deviation of all data when the data are explicitly separated into two separate populations is not a reasonable approach for Commerce’s Cohen’s *d* test.” Appx2417.

Finally, although both the simple average and weighted average will remain constant and will not reflect the difference in the means between the two groups, as explained above, Commerce determined that a simple average is a reasonable method of calculating the denominator of the Cohen’s *d* coefficient. Appx2410-2413.

D. PT’s Arguments That Commerce Should Use Weight Averaging Are Unavailing

PT’s attempts to demonstrate that Commerce should use a weighted average rather than a simple average to calculate the pooled standard deviation fail.

First, PT argues that Commerce has not explained how simple averaging is a reasonable methodology when applied to the facts of this investigation. Applnt. Br. 28. However, the facts of this investigation – that the groups consist of separate populations with unequal variances – has been at the center of Commerce’s explanation. Commerce explained that weighting is appropriate when the sample sizes of the groups are not equal because the group with a larger sample size has greater reliability and more weight is given to that group, and that the weights reflect the relative reliability of the standard deviation from the two groups. Appx2412. Because both groups in Commerce’s Cohen’s *d* test are populations that are equally reliable, Commerce uses a simple average. Appx2413. Moreover, Commerce based its explanation for using a simple average on evidence on the record. Reliability is a general statistical principle, and Commerce’s analysis of the sample size requirements for simple and weighted averages and reliability is supported by the academic literature on the record. Appx2410-13 (citing Appx527, Appx531).

Second, PT challenges the trial court’s findings that “{PT}’s examples serve to illustrate how weighting would work; they do not undermine the reasonableness of Commerce’s use of a simple average” and that the examples have “little analytical value.” Applnt. Br. 40 (citing Appx19). PT contends that five graphic examples demonstrate that weight averaging is reasonable and simple averaging is

not, and that low quantities create distortions. *Id.* at 28-40. However, the “only differences are the results themselves, and the arithmetic logic that different outcomes result when different weights are used to combine the standard deviations in the denominator of the Cohen’s d coefficient.” Appx2451-2452. To determine whether the differences in prices are significant, Commerce compares the Cohen’s d coefficient to the 0.8 threshold. Appx76. However, the graphical representations of hypothetical examples do not show whether a given set of compared prices represents a Cohen’s d coefficient larger or smaller than the 0.8 threshold. Appx2452. Therefore, the graphical representations do not express whether the data represented therein differ significantly. “In other words, there is no visual distinction between any of the graphical representations of the test and comparison group prices which would lead a reasonable observer to recognize that one difference in prices pass the Cohen’s d test and another difference in prices does not pass the Cohen’s d test, irrespective of whether a simple average or a weighted average is used.” *Id.*

The graphical representations only demonstrate an arithmetic tautology. As Commerce explained:

When the weights for averaging two values change from being identical (e.g., one) to being non-equal values, the results will change. When the weights are based on the sales quantities of each group, the smaller group will have less weight than the larger group, and the value being average (i.e., the standard deviation) will have a

smaller impact on the calculated average, and conversely the value of the larger group will have a larger impact. If the standard deviation of the smaller group is small, then the calculated average will be larger and the Cohen's *d* coefficient will be smaller. If the standard deviation for the smaller group is larger, then the calculated average will be smaller and the Cohen's *d* coefficient will be larger.

Appx2452-2453.

Third, PT asserts that methodology is validated by the results. Applnt. Br. 40-41. However, the results do not make a methodology reasonable. Appx2453. As Commerce explained, weighting is appropriate when the sample sizes of the groups of sampled data are not equal because the group with a larger sample size has greater reliability and more weight is given to that group. Appx2412. Thus, the weights reflect the relative reliability of the standard deviation from the two groups. Appx2412. Because both groups in commerce's Cohen's *d* test are populations that are equally reliable, Commerce uses a simple average. Appx2413. Moreover, as the trial court explained, "Commerce's use of Cohen's *d* in differential pricing calculations is not a visual analysis, but rather is a statistical methodology" and Commerce is tasked with developing a reasonable methodology, not a perfect one. *Mid Continent VII*, 680 F. Supp. at 1356 n.12.

Fourth, PT argues that administrative practice and judicial precedent support reliance on weighted averaging. Applnt. Br. 41-45. PT's argument that Commerce's differential pricing analysis is "substantially similar" to Commerce's

dumping margin calculation and administrative and court cases prefer weighted averages over simple averaging even in complex analysis is premised on a false equivalence. PT erroneously equates Commerce's Cohen's d test with the calculation of a respondent's weighted-average dumping margin. The calculation for Commerce's weighted-average dumping margin is not the Cohen's d test. Appx2454-2455. The weighted-average dumping margin has the total value of U.S. sales as the denominator. The statute defines the dumping margin as the "amount by which normal value exceeds the export price or constructed export price" and the weighted average dumping margin as the "percentage determined by dividing the aggregate dumping margins determined {} by the aggregate export prices and constructed export prices{.}" 19 U.S.C. § 1677(35)(A)-(B). Neither the calculation of individual dumping margins nor the aggregation of these dumping margins to calculate the weighted-average dumping margin involves the advanced statistical concepts at issue in this case, such as standard deviations or an examination of the significance of the difference between two groups.

Rather, Commerce uses the Cohen's d test to examine the difference in the mean prices between the test group and the comparison group and determines whether that difference is significant based on the variances of the prices within the two groups. Appx2455. PT disregards the purpose of the Cohen's d test. The purpose of the Cohen's d test is simply to examine whether prices in the U.S.

market differ significantly. Whereas a determination as to whether dumping exists requires both a U.S. price and a normal value (*i.e.*, home market price, third country price, or constructed value), Appx2455, the Cohen's *d* test does not examine "targeted dumping," includes no comparison with normal values, and only involves an analysis of U.S. prices. Appx76; Appx2454-2455. Thus, as the trial court found that PT "fail{s} to acknowledge that Commerce's task in its differential pricing methodology serves a diagnostic purpose" and that "Congress' grant of authority to Commerce dictates that diagnostic purpose." *Mid Continent VII*, 680 F. Supp. 3d at 1355 (citing Appx2455 and 19 U.S.C. § 1677f-1(d)(1)(B)). "Dumping margin calculations simply do not determine whether the difference in prices between the two groups is significant or 'the degree to which the phenomenon is present in the population,' but rather the potential uncollected dumping duty due.'" *Id.* (citations omitted).

Therefore, PT's reliance on judicial and administrative decisions involving the use of weight averaging in the context of determining individual dumping margins as supporting the use of weight averaging over simple averaging here is misplaced. Applnt. Br. 44-45. These decisions are also inapposite because they do not involve the advanced statistical concepts at issue in the Cohen's *d* test. Appx2454-2455.

Fifth, PT opposes the use of a simple average, arguing that it does not make

economic or statistical sense because lower volume sales, which PT describes as economically inconsequential, could be used to find whether prices differ significantly. Applnt. Br. 46. PT further argues that the test and comparison groups “do not have independent existences” and that each sale is a member of multiple test and comparison groups such that a particular sale will have more or less weight depending on which group the sale falls into and which group is designated as the test or comparison group. *Id.* at 47-48. PT’s arguments are misplaced.

The statute requires Commerce to examine whether “there is a pattern of export prices (or constructed export prices) for comparable merchandise that differ significantly among purchasers, regions, or periods of time.” 19 U.S.C. § 1677f-1(d)(1)(B)(i). To examine the pattern requirement, Commerce first uses the Cohen’s *d* test to determine whether the prices to a given purchaser, region, or time period differ significantly. Appx2405. In determining whether prices differ significantly, Commerce splits the U.S. sales into test and comparison groups. Appx2405, Appx2455. Each iteration of the Cohen’s *d* test represents a separate determination whether the prices to the test group differ significantly from the prices to the comparison group. Each iteration of the Cohen’s *d* test represents a different determination whether the prices to a separate test group differ significantly from the prices to a different comparison group. The “weight”

(whether by observation or sale quantity) of each sale price within each test group and comparison group is reflected in the weighted-average mean and weighted standard deviation of each group which is used to calculate the Cohen's d coefficient. Thus, the "weight" of each sale price is accounted for in Commerce's Cohen's d test. Appx2414, Appx2455.

Under Commerce's Cohen's d test, a sale will fall into a test group once for each of the three categories tested, *i.e.*, by purchaser, by region, and by time period. Further, each sale will fall into one or more comparison group that will depend on the number of sales to other purchasers, in other regions, and during other time periods. PT states that "a particular sale will receive more weight than other sales for certain comparisons . . . while in other comparisons the same sale will receive a low weight." Applnt. Br. 48. However, in each test and comparison group, each sale is weighted when calculating the mean and standard deviation of each test and comparison group. Appx2455. Thus, the "weights" of each sale, *i.e.*, the quantity sold, are reflected in the mean and standard deviation of the prices within each group.

Sixth, PT asserts that Commerce's rationale for rejecting the single standard deviation is also why a simple average fails, asserting that ANOVA does not support simple averaging and a single standard deviation. Applnt. Br. 56-62. However, PT fails to explain how an ANOVA analysis is relevant to Commerce's

analysis of the difference in the mean prices, either based on the academic literature or otherwise. Appx2459. Finally, ANOVA is used with Dr. Cohen's f coefficient, which "is used to measure the dispersion of means among three or more groups" and a distinct concept from Cohen's d .⁴ Appx498; Appx2459. Therefore, this contention should fail.

III. The New Issues And Arguments Raised By Canada Should Be Rejected Because They Stray Beyond The Sole Question Before The Court And The Arguments Raised In PT's Opening Brief And Are Supported By Non-Record Evidence

The only question before the Court in this appeal is whether Commerce's use of a simple average in the denominator of the Cohen's d coefficient is reasonable. *See Mid Continent V*, 31 F.4th at 1369-70; *Mid Continent VII*, 680 F. Supp 3d at 1351 ("Only the Cohen's d test, which determines whether there is a 'pattern of prices that differ significantly,' is at issue in this case.") (citations omitted); Applnt. Br. 3 ("The specific issue for this Court is whether Commerce's

⁴ PT refers to factual information that it had improperly included as Attachment 1 to the Reply Brief of Defendants-Appellants, *Mid Continent Steel & Wire Inc. v. United States*, Case No. 2021-1747 (Fed. Cir. Sept. 3, 2021), ECF No. 23. Applnt. Br. 17 n.7. Attachment 1 to PT's reply brief in *Mid Continent V* consists of pages 359-61 of Dr. Cohen's textbook, and it was not a part of the record in this proceeding when it was filed with the Court. Commerce subsequently rejected these pages, and the trial court sustained Commerce's rejection of them. *Mid Continent VI*, 628 F. Supp. 3d at 1319-20. Although these pages were referenced by the Court in *Mid Continent V*, they were inappropriately placed before the Court by PT. This sleight of hand should not be permitted to continue, and these documents should not be considered by the Court in this appeal.

calculation of the standard deviation denominator based on the simple average ('SA') of the {standard deviations} of the {weighted average} prices in each subgroup is supported by substantial evidence and in accordance with law.”). However, Canada has raised new issues that were not raised by PT during the remand proceedings, before the trial court, or in PT’s opening brief in this appeal. These new issues should be categorically rejected.

Canada first questions whether Dr. Cohen’s thresholds can be used when “the data violate the assumptions of equal variance and equal size.” Amici Br. 27-30. Canada argues that, regardless of the rationale Commerce uses for simple averaging the denominator when the groups are of different sizes, the resulting d coefficient cannot be interpreted using the “small, medium, and large thresholds” because Dr. Cohen created those thresholds using data that contextualizes mean differences in a single, common scale. *Id.* at 30. Canada also questions whether Commerce selected the correct formula of the four alternative formulas (referred to as “Methods”) used to calculate a single standard deviation described by the academic literature. *Id.* at 30-34. GOC argues that Commerce used the wrong “Method” when it determined that a single standard deviation was not an appropriate denominator in the Cohen’s d test. *Id.* at 30, 32. Although we recognize that the Court has discretion to consider new arguments on appeal, *Mobility Workx, LLC v. Unified Patents, LLC*, 15 F. 4th 1146, 1151 (Fed. Cir.

2021), for several reasons, the Court should reject GOC’s attempt to present these issues and the arguments supporting them.

A. Canada’s New Issues And Arguments Should Be Deemed Waived Because PT’s Brief Does Not Raise Or Otherwise Include Them

This is PT’s appeal, it is not a joint appeal by PT and Canada, and PT did not raise these issues in its opening brief. *See Amoco Oil Co. v. United States*, 234 F.3d 1374, 1378 (Fed. Cir. 2000) (“{i}t is the appellant’s case, not a joint appeal by the appellant and *amicus*. Appellant must raise in its opening brief all the issues it wishes the court to address.”); *Sage Prods., Inc. v. Devon Indus., Inc.*, 126 F.3d 1420, 1426 (Fed. Cir. 1997) (“With a few notable exceptions, such as some jurisdictional matters, appellate courts do not consider a party’s new theories, lodged first on appeal.”); *Nan Ya Plastics Corp. v. United States*, 810 F.3d 1333, 1350 (Fed. Cir. 2016) (holding that contentions based on statistical methodologies never raised before Commerce or the trial court were waived). Canada’s new issues are not jurisdictional.

Canada’s arguments regarding thresholds are not appropriately made here. Amici Br. 27-30. In addition to the fact that PT does not make a thresholds argument here, although PT had originally challenged Commerce’s use of Dr. Cohen’s thresholds to determine whether prices differed significantly, the Court rejected PT’s argument and affirmed Commerce’s use of the thresholds as reasonable. *Mid Continent III*, 940 F.3d at 673, 674-675. The decision in *Mid*

Continent III on the thresholds issue constitutes the law of this case. *See Koyo Seiko Co. v. United States*, 95 F.3d 1094, 1097 (Fed. Cir. 1996) (citing *Arizona v. California*, 460 U.S. 605, 618 (1983) (“{T}he {law of the case} doctrine posits that when a court decides upon a rule of law, that decision should continue to govern the same issues in subsequent stages in the same case.”); *Robinson v. McDonough*, No. 2021-1784, 2022 WL 499845, at *2 (Fed. Cir. Feb. 18, 2022) (“Under the law-of-the-case doctrine, courts generally refuse to reconsider questions of law and fact that have already been decided during litigation to ‘prevent relitigation of issues.’”) (quoting *Suel v. Sec’y of Health & Hum. Servs.*, 192 F.3d 981, 984–85 (Fed. Cir. 1999)). Canada impermissibly attempts to relitigate an issue decided in an earlier stage of this case and, therefore, Canada’s thresholds argument should be rejected.

Likewise, Canada’s “Methods” argument should be rejected as waived. A party must also develop an argument in its opening brief. *See SmithKline Beecham Corp. v. Apotex Corp.*, 439 F.3d 1312, 1320 (Fed. Cir. 2006) (holding that a party must articulate a “developed argument” in its opening brief to avoid waiver). PT’s brief discusses the use of a single standard deviation, its arguments are substantially different than Canada’s arguments, and it has not developed the argument the Canada makes. Indeed, in its comments to Commerce during the fourth remand segment, PT ***agreed*** that Commerce correctly concluded that the use

of a single standard deviation is not supported by the academic literature for use as the denominator in the Cohen's d coefficient. Appx2432, Appx2435. However, even though PT argues now that the test and comparison groups are not independent groups and relying on the single standard deviation is consistent with weighted averaging, Applnt. Br. 55-56, PT never pressed or even remotely developed the argument that Canada makes that Commerce used the wrong "Method" when determining that a single standard deviation is not an appropriate denominator in Commerce's Cohen's d coefficient. Because PT did not adopt or develop the argument that Canada makes in its brief and Canada's arguments are not subsumed within PT's arguments, this argument should be deemed waived.

B. Canada's New Arguments And Non-Record Supporting Materials Should Be Stricken Or Disregarded Because They Were Not Before Commerce During The Administrative Proceedings

Canada's thresholds and standard deviation arguments were not raised by interested parties in the underlying administrative segment before Commerce. Therefore, Commerce did not have an opportunity to address these arguments in the fourth remand redetermination. Commerce's determination cannot be fairly reviewed for error on issues and arguments that were never raised to Commerce during its administrative segment in this proceeding and, consequently, not included in its decision. *United States v. L.A. Tucker Truck Lines, Inc.*, 344 U.S. 33, 37 (1952) ("Simple fairness to those who are engaged in the tasks of

administration, and to litigants, requires as a general rule that courts should not topple over administrative decisions unless the administrative body not only has erred but has erred against objection made at the time appropriate under its practice.”); *Corus Staal BV v. United States*, 502 F.3d 1370, 1379 (Fed. Cir. 2007) (“Absent a strong contrary reason, the court should insist that parties exhaust their remedies before the pertinent administrative agencies.”). Because Commerce did not have an opportunity to address these specific issues and the arguments supporting them, the Court should not consider these arguments now when reviewing Commerce’s fourth remand redetermination.

Relatedly, Canada supports its new arguments with factual information that is not part of the administrative record and, therefore, was not considered by Commerce in this proceeding. Congress has determined that Commerce must base its determinations upon the record developed during the administrative proceeding. 19 U.S.C. §§ 1516a(b)(1) & (2). The statute also establishes the standard for judicial review of Commerce’s determination, and review is based in part on the “evidence on the record.” 19 U.S.C. § 1516a(b)(1)(B)(i). The U.S. Supreme Court has explained that “the focal point for judicial review should be the administrative record already in existence, not some new record made initially in the reviewing court.” *Camp v. Pitts*, 411 U.S. 138, 142 (1973). *See also Home Prods. Int’l, Inc. v. United States*, 633 F.3d 1369, 1379 (Fed. Cir. 2011). This Court has further

explained that “{t}he purpose of limiting review to the record actually before the agency is to guard against courts using new evidence to ‘convert the “arbitrary and capricious” standard into effectively de novo review.’” *Axiom Res. Mgmt., Inc. v. United States*, 564 F.3d 1374, 1380 (Fed. Cir. 2009) (quoting *Murakami v. United States*, 46 Fed. Cl. 731, 735 (2000), *aff’d*, 398 F.3d 1342 (Fed. Cir. 2005)).

Finally, the question of whether a methodology is reasonable must be evaluated within the confines of information that was on the administrative record at the time the agency made its decision. *Cf. Ad Hoc Shrimp Trade Action Comm. v. United States*, 791 F. Supp. 2d 1327, 1335 (Ct. Int’l Trade 2011) (“As the antidumping statute is silent on this particular question, the Court will uphold Commerce’s reasonable methodology *if it comports with reasonable reading of the administrative record.*”) (emphasis added).

Academic literature and data are factual information that must be submitted to the record of the administrative proceeding for Commerce to consider. 19 C.F.R. § 351.102(b)(21) (defining “factual information” to include “evidence, including statements of fact, documents, and data”). *See also Stupp*, 5 F.4th at 1348-51 (Fed. Cir. 2021) (concluding that Commerce had properly rejected non-record academic text); *Mid Continent VI*, 628 F. Supp. 3d at 1319-20 (sustaining Commerce’s rejection of an untimely expert report and academic text); *Tri Union Frozen Prod., Inc. v. United States*, 163 F. Supp. 3d 1255, 1290 (Ct. Int’l Trade

2016) (explaining that an academic paper is factual information that must be placed on the record within the applicable time limits).

Because Canada relies on materials that are not part of the administrative record and were never presented to Commerce either during the investigation or on remand, the Court should strike or disregard these non-record materials and any arguments that rely on them. These materials include:

- Larry V. Hedges, *Review and Analysis of the Cohen's d Test as Used in the U.S. Department of Commerce's Differential Pricing Methodology* (Jan. 2, 2024).
- Steven Wachs, *What Is a Standard Deviation and How Do I Compute It?* at 3 (2009).
- J.P. Guilford, *Fundamental Statistics in Psychology and Education* at 51–52 (1st ed. 1942).
- R.E. Deakin, D.G. Kildea, *A Note on Standard Deviation and RMS*, 44 *The Australian Surveyor* 74 (June 1999).
- Johnson Ching-Hong Li, *Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data*, 48 *Behavioral Research* 1560, 1561 (2016).
- Xiang Wan, Wenqian Wang, Jiming Liu, Tiejun Tong, *Estimating the Sample Mean and Standard Deviation from the Sample Size, Median, Range and/or Interquartile Range*, 14:135 *BioMed Central Medical Research Methodology* at 2 (2014).
- Jum C. Nunnally, *Psychometric Theory — 25 Years Ago and Now* at 10 (Nov. 1975).

- CDC, Data Table of Stature-for-age Charts, available at https://www.cdc.gov/growthcharts/html_charts/statage.htm#females (last updated Aug. 23, 2001).
- Cohen at 535-537.

Canada understands that it has introduced non-record factual information into this appeal, and it requests that the Court take judicial notice of this material pursuant to Federal Rule of Evidence 201(b). Amici Br. 29 n.89. In addition to the reasons stated above, the Court should decline to take judicial notice of the non-record factual information because “{j}udicial notice is not meant to circumvent the creation and review of an agency record.” *Tri Union Frozen Prods. v. United States*, 161 F. Supp. 3d 1333, 1340 (Ct. Int’l Trade 2016).

C. Canada’s New Arguments Are Unavailing

Even if the Court were to consider Canada’s new arguments, which it should not, Canada’s new arguments fail.

First, Canada erroneously asserts that there is one population, *i.e.*, a “broader group,” whose standard deviation should be the denominator in Commerce’s Cohen’s *d* test. Amici Br. 12 n.38, 33. This assertion is a distortion of the academic literature, the Cohen’s *d* test, and the statutory framework of the pattern requirement. Indeed, Canada ignores Dr. Cohen’s general formulation of the effect size, where the denominator, σ , is defined as *either* the standard deviation of either population A or the standard deviation of population B when the standard

deviation of population A is assumed to be equal to the standard deviation of population B. Appx2414 (citing Appx534 and Appx537). Therefore, as Commerce correctly determined, “the denominator, σ , in Dr. Cohen’s equations 2.2.1 and 2.2.2, is either the standard deviation of population A or the standard deviation of population B, but it is **not** the standard deviation of populations A and B combined together.” Appx2414. Simply put, Dr. Cohen’s denominator is not the standard deviation of *commingled* populations A and B. This same recognition of two separate populations continues in Dr. Cohen’s equation 2.3.2, where the alternative denominator, σ' , is the “root mean square” of the standard deviations of the two populations. Appx546. Nowhere does Dr. Cohen contemplate the use of the standard deviation of commingled data from populations A and B.

Furthermore, Commerce’s Cohen’s *d* test uses the two independent groups designs.⁵ Appx2414, Appx2455. In general, the differential pricing analysis considers whether the A-to-A method is appropriate to calculate the respondent’s weighted-average dumping margin, and, thus, the Cohen’s *d* test is limited to the respondent’s U.S. prices of subject merchandise during the period of investigation or review. Appx75-76. Although the respondent may sell many different products, the pattern requirement focuses on the prices of comparable merchandise

⁵ Goulet-Pelletier & Cousineau describe four scenarios in which Cohen’s *d* may be used including “in two independent groups designs, where the interest is in the difference between two population means.” Appx710.

to a given purchaser, region, or time period (the test group) in comparison with the prices to all other purchasers, regions or time periods (the comparison group). 19 U.S.C. § 1677f-1(d)(1)(B)(i). Thus, the statutory framework, as reflected in the Cohen's d test, creates two independent groups, which are also independent populations, *i.e.*, they include all prices of relevant U.S. sales. The test and comparison groups are not taken from a single "broader group" where the change in the mean price is measured, as posited by Canada. The prices in each group have not been sampled from some larger population or populations, or some "broader group," as alleged by Canada. The Cohen's d test measures the difference in the mean prices between two independent and distinct groups. Appx2455.

Dr. Cohen's formulation of the d coefficient also follows the same two independent groups designs. Appx2414. In general, Dr. Cohen's discussion presents the d coefficient as measuring the effect size between two populations. Appx534. Further, in the real-world observations, Dr. Cohen illustrates his proposed small, medium, and large effect thresholds using the two independent group design. For example, Dr. Cohen compares the heights of teenage girls of various ages. Appx536-537. Dr. Cohen finds that the difference in the height of 13- and 18-year-old girls demonstrates a large, "grossly perceptible" difference. Appx537. Although "13- and 18-year-old girls" are both girls and both teenagers,

that does not mean that these two groups belong to a “broader group” from which these two groups have been sampled. Combining 13- and 18-year-old girls into a broader group would be mixing apples and oranges into a single “population,” to which one could also include other ages, as well as boys. Indeed, Dr. Cohen gave such an example using males and females of a “research plan which randomly samples the two populations (simple randomized design or comparison between two independent means).” Appx530.

To support its argument, Canada relies on Jean-Christophe Goulet-Pelletier & Denis Cousineau, *A Review of Effect Sizes and Their Confidence Intervals, Part I: The Cohen’s d Family*, 14 *The Quantitative Methods for Psychology*, 242, 243 (2014) (Appx2021-2022) (Goulet-Pelletier & Cousineau). Amici Br. 13 & n.41, 14 & n.44. However, Goulet-Pelletier & Cousineau is consistent with Cohen, stating that for a two independent group design, “{a}ssuming that both groups are from *populations* with the same variance, the best estimate of the population standard deviation σ is a weighted average of the standard deviations of both samples.” Appx716 (emphasis added). Here, Goulet-Pelletier & Cousineau estimates the standard deviation of population A *or* the standard deviation of population B, which are assumed to be equal, based on the weighted average of the standard deviations of both samples. *See* Appx557 and equation 2.5.2. Goulet-Pelletier & Cousineau do not use a single population standard deviation.

Therefore, Canada's reliance on Goulet-Pelletier & Cousineau is misplaced and fails to support Canada's arguments.

Canada's reliance on Coe as supporting its single standard deviation argument, Amici Br. 21, is also misplaced. Canada wholly ignores Professor Coe's warning that "{t}he use of a pooled estimate of standard deviation depends on the assumption that the two calculated standard deviations are estimates of *the same* population value. . . . Where this assumption cannot be made (either because there is some reason to believe that the standard deviations are likely to be systematically different, or if the actual measured values are very different), then a pooled estimate should not be used." Appx568 (emphasis in original). Coe's warning would be unnecessary if there is only one population because the assumption would never fail to be "the same population value." Thus, there is more than one population, and the only reason there is one population standard deviation is because the two populations are assumed to have the same variance. Appx534.

Second, Canada's argument that the single standard deviation critiqued by Commerce is not the "Method" prescribed by the literature, Amici Br. 30-32, is unavailing. Canada fails to explain how the critiques of this formula do not apply to populations. Professor Coe's warning that the single standard deviation would be partly determined by the difference in the means of the two groups holds true

for populations just as it does for sampled data. Therefore, if Canada's statement that Method 1 is used for populations and Method 4 is used for samples is correct, a single standard deviation calculated using Method 1 based on populations or Method 4 based on sampled data would suffer from the same flaw. That is, the resulting denominator would include not only the variances within the two groups, but also reflect the difference in the means between the two groups.

Third, Canada also mistakes $\bar{X}_{E\&C}$ to be the means of the test and comparison group rather than the mean of the commingled group of both the test and comparison group, or in other words, the mean of all the U.S. sales. Amici Br. 17; Appx2417. Indeed, in the examples in the appendix of the remand, Commerce (and PT) used the mean of all the U.S. sales for the variance and standard deviation of all the U.S. sales. Appx2465-2467 (defining the mean as the total value over the total quantity); Appx1792-1797 (stating the same and that stating that the examples are populations). These examples show that the critique stated in the remand redetermination is applicable even to populations, as with sampled data, and that the standard deviation of all U.S. sales would vary with the value of the difference in the means, which is not true of the weighted average or simple average. Appx2415-2416, Appx2442, Appx2465-2467.

Finally, although Canada's arguments regarding thresholds, Amici Br. 27-30, are not appropriately made here, as we discuss above, they are also unavailing.

Contrary to GOC’s assertion, Dr. Cohen’s thresholds are not based on a statistical analysis, and Dr. Cohen has described the thresholds as arbitrary, explaining that “{a}lthough arbitrary, the proposed conventions will be found to be reasonable by reasonable people.” Appx530.

CONCLUSION

For all of these reasons, we respectfully request that the Court affirm the trial court’s judgment and sustain Commerce’s remand redetermination.

Respectfully Submitted,

BRIAN M. BOYNTON
Principal Deputy Assistant Attorney General

/s/Patricia M. McCarthy
PATRICIA M. MCCARTHY
Director

/s/ Mikki Cottet
MIKKI COTTET
Senior Trial Counsel
Commercial Litigation Branch
Civil Division
U.S. Department of Justice
Classification Unit, 8th Floor
P.O. Box 480, Ben Franklin Station
Washington, D.C. 20044
Tel: (202) 307-0962
Fax: (202) 514-7965

OF COUNSEL:
VANIA WANG
Attorney
U.S. Department of Commerce
Office of the Chief Counsel for
Trade Enforcement & Compliance
Washington, D.C. 20230

August 15, 2024

Attorneys for Defendant-Appellee

CERTIFICATE OF COMPLIANCE

1. I hereby certify that the foregoing brief complies with the Rules of this Court in that it contains 13,939 words including text, footnotes, and headings. This is within the limit of 14,000 words permitted under Federal Circuit Rule 28.1, excluding the parts of the brief exempted by Federal Rule of Appellate Procedure 32(f) and Federal Circuit Rule 32(b).

2. This brief complies with the typeface requirements of Federal Rule of Appellate Procedure 32(a)(5) and the type style requirements of Federal Rule of Appellate Procedure 32(a)(6). The brief has been prepared in a proportionally spaced typeface using Microsoft Word in 14-point Times New Roman font.

/s/ Mikki Cottet

CERTIFICATE OF SERVICE

I hereby certify under penalty of perjury that on this 15th day of August, 2024, a copy of the foregoing brief

X was filed electronically.

X This filing was served electronically to all parties by operation of the Court's electronic filing system.

/s/ Mikki Cottet